

Universidade Técnica de Lisboa
Instituto Superior de Economia e Gestão

Mestrado em Matemática Financeira

Trabalho Final de Mestrado

CLUSTERS DE SÉRIES FINANCEIRAS

Filipa Martins Parrinha

Setembro de 2012

Universidade Técnica de Lisboa
Instituto Superior de Economia e Gestão

Mestrado em Matemática Financeira

Trabalho Final de Mestrado

CLUSTERS DE SÉRIES FINANCEIRAS

Por Filipa Martins Parrinha

Orientação:

Prof. Doutor Jorge Caiado

Setembro de 2012

Conteúdo

1	Agradecimentos	3
2	Introdução	4
3	Análise de séries temporais financeiras	6
3.1	Propriedades das Séries Financeiras	6
3.2	Modelos Heterocedásticos	8
4	Metodologia	13
4.1	Classificação e agrupamento de séries temporais	13
4.2	Medidas de distância	16
5	Simulação	20
6	Aplicação	24
6.1	Informação Geral	24
6.2	Estatísticas Descritivas	25
6.3	Estimação GARCH	25
6.4	Análise de Clusters	26
6.4.1	Dependência de curto prazo	27
6.4.2	Dependência de longo prazo	30
6.4.3	Discussão	34
7	Conclusão	35

A	Gráficos	40
B	Empresas constituintes do índice FTSE100 a 1 de Maio de 2012	42
C	Estatísticas descritivas	46

Capítulo 1

Agradecimentos

Agradeço primeiramente ao professor Dr. Jorge Caiado pelo tempo dedicado, apoio, conhecimento e orientação que investiu neste projeto.

Agradeço também aos meus pais, José Jorge Martins e Maria Filomena Parrinha, por todo o apoio, educação, dedicação, amizade e compreensão, e à minha irmã, Beatriz Martins, pela eterna paciência e dedicação.

Ao meu namorado, Nuno Fernandes, agradeço-lhe a enorme dedicação e amizade que me ofereceu ao longo deste projeto.

Por último, mas não com menos importância, gostaria de deixar um enorme agradecimento a todos os meus amigos pela sua enorme amizade.

Capítulo 2

Introdução

Desde as décadas de 60/70 que, fruto da globalização dos mercados financeiros, se assiste a uma diversificação de instrumentos financeiros bem como a um aumento de agentes financeiros, de liquidez e volume de transações. Estes acontecimentos tornaram os mercados financeiros mais complexos e consequentemente o estudo das séries financeiras uma área de bastante interesse. Por outro lado, a crescente facilidade de acesso a dados financeiros, em número elevado e com maior frequência (diária, mensal, anual), contribuiu para a evolução deste estudo nos últimos anos (ver [Nicolau, 2008]).

A partir de 1900, com a introdução do movimento Browniano nas finanças através da tese *La Théorie de la spéculation* do francês Bachelier, foram surgindo modelos para tentar ajustar o comportamento dos retornos e da variância condicional (volatilidade) de ativos financeiros, desde modelos contínuos a modelos discretos (ver [Engle, 1982]). Estes modelos têm bastante relevância na atividade financeira pois são indispensáveis para a determinação do preço justo ou prêmio de uma opção, modelação e previsão da volatilidade (necessária na gestão de risco, na afetação e seleção de ativos financeiros, na valorização e *hedging* de derivados financeiros), gestão e constituição de portfólios.

A análise dos preços diários dos ativos financeiros revelou comportamentos e

características estatísticas comuns (ver [Mandelbrot, 1963]), motivando o interesse da identificação de semelhanças entre séries financeiras para efeitos de diversificação de carteiras de ativos e controlo de risco. Assim, foram-se desenvolvendo técnicas de análise multivariada para classificar e agrupar estas séries temporais.

Pretende-se com este trabalho estudar semelhanças entre séries financeiras, usando métodos de análise multivariada baseados em modelos discretos. Um método com bastante destaque é o método de aglomeração (*clustering*) que pressupõe a escolha adequada de uma medida de distância entre séries temporais. A medida de distância clássica é a medida euclidiana entre os respectivos pontos no espaço, no entanto esta medida apresenta três grandes limitações (ver [Galeano & Peña, 2000]). Como alternativa propõe-se a utilização da medida de distância entre as respetivas funções de autocorrelação dos módulos dos retornos, em detrimento do quadrado dos retornos, como foi introduzido por Caiado e Crato (2010). Esta distância permite captar a proximidade entre as volatilidades de duas séries temporais e, conseqüentemente, analisar comportamentos semelhantes entre séries financeiras.

Este trabalho encontra-se estruturado da seguinte forma: o capítulo 2 apresenta alguns factos estilizados presentes nas séries financeiras, assim como alguns dos modelos heterocedásticos mais frequentes em finanças; no capítulo 3 mencionam-se algumas técnicas estatísticas de análise multivariada e medidas relevantes na aplicação das mesmas; o capítulo 4 conduz uma simulação que compara o desempenho de duas medidas usando métodos de agrupamento de séries financeiras; no capítulo 5 desenvolve-se um estudo de *clustering* aplicado ao índice FTSE100.

Capítulo 3

Análise de séries temporais financeiras

3.1 Propriedades das Séries Financeiras

A generalidade das séries financeiras exibem certas características estatísticas comuns (factos estilizados). Uma parte delas foi inicialmente apresentada por Mandelbrot em 1963, e depois completada por outros autores. Alguns dos factos estilizados são os seguintes (ver [Cont, 2001]):

1. **Ausência de autocorrelações:** os retornos dos ativos apresentam autocorrelações lineares insignificantes (nulas ou quase nulas), assim como as séries temporais dos resíduos;
2. **Caudas pesadas:** os retornos apresentam distribuições leptocúrticas, pois as observações extremas (retornos muito altos ou muito baixos) ocorrem com maior frequência do que se a sua distribuição fosse normal;
3. **Ganhos e perdas assimétricos:** as descidas de preços observados tendem a ser superiores às subidas de preços, obtendo-se assim uma representação assimétrica dos retornos dos ativos;

4. ***Aggregational Gaussianity***: o aumento da escala temporal (Δt) ou diminuição da frequência das observações (por exemplo de observações diárias para semanais ou mensais) no cálculo dos retornos faz com que a distribuição destes se aproxime mais da distribuição Normal;
5. **Não estacionaridade**: as trajetórias dos preços são semelhantes ao passeio aleatório;
6. ***Volatility Clustering***: valores extremos de retornos tendem a ocorrer de forma seguida, observando-se volatilidades muito altas por largos períodos de tempo; formam-se, assim, aglomerações (clusters) de valores de retornos muito altos e muito baixos e, consequentemente, clusters de volatilidades muito altas;
7. ***Conditional heavy tails***: as distribuições condicionais das séries temporais residuais exibem caudas pesadas, mas não tão pesadas como as das distribuições dos retornos;
8. **Forte dependência temporal da volatilidade**: a função de autocorrelação do módulo dos retornos exhibe um decréscimo lento, o que significa que se a volatilidade é alta (baixa) então é razoável esperar que a volatilidade se mantenha alta (baixa) durante bastante tempo, esta propriedade também é designada por memória longa;
9. ***Leverage Effect***: a volatilidade apresenta-se assimétrica relativamente às observações passadas positivas e negativas. Retornos negativos (correspondentes a descidas de preços) tendem a aumentar a volatilidade com uma amplitude maior do que os retornos positivos (subidas de preços) da mesma magnitude.

A complexidade do estudo dos mercados financeiros deve-se, em parte, ao facto de estas características serem difíceis de reproduzir usando processos estocásticos.

3.2 Modelos Heterocedásticos

De forma a satisfazer as necessidades da atividade financeira (controlo de risco, avaliação de instrumentos financeiros, previsão, gestão de portfolios, *hedging*, etc.) e dadas as características das séries financeiras, surge a necessidade de modelar a volatilidade (variabilidade dos retornos).

Modelos de heterocedasticidade são modelos que permitem estudar o comportamento da variância como sendo uma variável aleatória ao longo do tempo.

Seja $u_t = \sigma_t \epsilon_t$ tal que:

- $\{\epsilon_t\}$ é uma sucessão de variáveis aleatórias (v.a.) independentes e identicamente distribuídas (i.i.d.) com $E[\epsilon_t] = 0$ e $Var[\epsilon_t] = 1$, ou seja, $\{\epsilon_t\}$ é um ruído branco;
- $\{\epsilon_t\}$ é independente de u_{t-k} com $k \in \mathbb{N}$;
- σ_t é F_{t-1} mensurável, isto é, σ_t depende apenas de variáveis observadas até ao momento $t - 1$.

Assim tem-se que

$$E[u_t|F_t] = 0 \quad \text{e} \quad Var[u_t|F_t] = \sigma_t^2.$$

O processo u_t diz-se um processo heterocedástico.

Os modelos de heterocedasticidade permitem a modelação da variância como variável aleatória, em vez de considerá-la como constante ao longo do tempo.

Os mais conhecidos são os modelos de heterocedasticidade condicionada que modelam a volatilidade (variância condicionada à informação anterior fornecida pelo próprio processo) de uma série temporal. O primeiro modelo que surgiu foi apresentado por Engle em 1982, o modelo autoregressivo de heterocedasticidade condicional, ARCH, que declara que a variabilidade do quadrado da variância pode

ser modelada por uma média móvel de observações passadas da série temporal (ver [Engle, 1982]):

$$u_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2$$

onde os parâmetros α_i satisfazem as seguintes condições:

- $\alpha_0 > 0$
- $\alpha_i \geq 0, \quad i = 1, 2, \dots, q$

de forma a assegurar que a variância não condicionada seja positiva. Neste caso diz-se que u_t segue um modelo ARCH de ordem q , ARCH(q).

Os modelos ARCH permitem que o processo u_t tenha uma distribuição marginal leptocúrtica e com cauda pesada, tornando-os muito populares na análise de séries financeiras. Adicionalmente são fáceis de lidar, admitem *volatility clustering* devido à sua capacidade de captar a dependência temporal da variância condicionada e são não lineares. No entanto, como consequência da forte dependência temporal da volatilidade, seria necessário o uso de ordens muito elevadas para ajustar o modelo a séries financeiras, o que o tornaria não tratável.

Como forma de superar esta limitação, Bollerslev introduziu o modelo autoregressivo de heterocedasticidade condicional generalizado, GARCH, (ver [Bollerslev, 1986]) onde a variância condicional é expressa em função dos quadrados dos seus valores passados e dos quadrados dos valores passados dos erros da série temporal:

$$u_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2$$

onde ϵ_t são v.a. i.i.d., com média nula e variância unitária, independentes das variações passadas de u_t . Para garantir que σ^2 seja positiva tem-se que $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, 2, \dots, q$ e $\beta_j \geq 0, j = 1, 2, \dots, p$.

Diz-se, neste caso, que u_t segue um modelo GARCH de ordem (p, q) , $\text{GARCH}(p, q)$. Note-se que no caso $p = 0$ o processo u_t reduz-se a um modelo $\text{ARCH}(q)$.

Tem-se assim uma estrutura de parâmetros mais parcimoniosa. Em particular, o modelo simples $\text{GARCH}(1, 1)$ tem-se revelado suficiente para captar a volatilidade na maioria das aplicações com séries financeiras. Esta versão do modelo não permite no entanto captar choques assimétricos na volatilidade, propriedade verificada em muitas séries financeiras.

Para ultrapassar esta dificuldade Nelson (1991) introduziu o modelo GARCH exponencial (EGARCH), definindo uma função de erros ponderados (ver [Tsay, 2005])

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)]$$

$$\Leftrightarrow g(\epsilon_t) = \begin{cases} (\theta + \gamma)\epsilon_t - \gamma E(|\epsilon_t|), & \text{se } \epsilon_t \geq 0 \\ (\theta - \gamma)\epsilon_t - \gamma E(|\epsilon_t|), & \text{se } \epsilon_t < 0 \end{cases}$$

onde θ e γ são constantes reais.

Desta forma, um modelo $\text{EGARCH}(p, q)$ é definido por

$$u_t = \sigma_t \epsilon_t,$$

$$\ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \dots + \beta_{q-1} B^{q-1}}{1 - \alpha_1 B - \dots - \alpha_p B^p} g(\epsilon_{t-1})$$

onde α_0 é uma constante, B é um operador tal que $Bg(\epsilon_t) = g(\epsilon_{t-1})$, e $1 + \beta_1 B + \dots + \beta_{q-1} B^{q-1}$ e $1 - \alpha_1 B - \dots - \alpha_p B^p$ são polinómios sem fatores comuns e cujos zeros em valor absoluto são maiores do que a unidade.

Assim, como consequência da propriedade assimétrica da função $g(\epsilon_t)$, este modelo permite efeitos assimétricos entre retornos positivos e negativos, captando o chamado *leverage effect* (ver [Black, 1976]) presente em muitas séries financeiras e

tornando-se uma alternativa ao modelo GARCH na modelação destas séries temporais.

Definindo a variância condicional como uma função das partes positivas e negativas dos seus valores passados também é possível introduzir assimetria no modelo GARCH (ver [Glosten et al., 1993] e [Zakoian & Rabemananjara, 1993]):

$$u_t = u_t^- + u_t^+, \quad \text{onde} \quad u_t^- = \min \{u_t, 0\}, \quad u_t^+ = \max \{u_t, 0\}.$$

Assim, u_t segue um modelo TGARCH(p, q) (*Threshold GARCH* de ordem p e q) se for da forma

$$\begin{aligned} u_t = \sigma_t \epsilon_t, \quad \sigma_t &= \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^- + \sum_{i=1}^q \gamma_i u_{t-i}^+ + \sum_{j=1}^p \beta_j \sigma_{t-j} \\ \Rightarrow \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{i=1}^q \gamma_i d_{t-1} u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \end{aligned}$$

onde $\{\epsilon_t\}$ é uma sequência de v.a. i.i.d., com média zero e variância unitária, independentes das variações passadas de u_t , $d_t = 1$ se $u_t < 0$ e $d_t = 0$ caso contrário.

Desta forma, este modelo também permite captar o *leverage effect*. Em particular, pode-se definir o modelo TGARCH(1,1) da seguinte forma

$$u_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha u_{t-1}^2 + \gamma u_{t-1}^2 d_{t-1}$$

onde $\{\epsilon_t\}$ é uma sequência de v.a. i.i.d., com média zero e variância unitária. Dependendo do valor de γ , os choques negativos ou a chegada de “más notícias” ($u_{t-1} < 0$) poderão fazer com que a volatilidade diminua ($\gamma < 0$), aumente ($\gamma > 0$) ou não se altere ($\gamma = 0$). As “boas notícias” terão um impacto de α na volatilidade, enquanto as “más notícias” terão um impacto de $\alpha + \gamma$. A persistência dos choques na volatilidade é dada por $\alpha + \beta + \gamma/2$.

Em finanças, o retorno de um ativo financeiro e a sua volatilidade estão relacionados, na medida em que um ativo financeiro com maior retorno terá tendencialmente uma maior volatilidade. Agentes económicos avessos ao risco devem receber uma compensação, o prémio de risco, por investirem em ativos financeiros com risco elevado. Quando se usam modelos ARMA-GARCH não é possível captar este comportamento, pois a média condicionada e a variância são modeladas de forma independente. Em 1987, Engle, Lilien e Robins apresentaram o modelo GARCH-M (GARCH *in the mean* ou GARCH na média) da seguinte forma (ver [Tsay, 2005] e [Francq & Zakoian, 2010])

$$r_t = \mu + c\sigma_t^2 + u_t, \quad u_t = \sigma_t\epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2$$

onde μ e c são constantes. O parâmetro c representa o prémio de risco. No caso de séries financeiras espera-se que $c > 0$, isto é, que o retorno e a sua volatilidade sejam positivamente relacionados.

Outros modelos econométricos têm sido desenvolvidos na literatura de análise de séries financeiras para descrever a variância condicional, como o modelo CHARMA (*conditional heteroscedastic ARMA*) (ver [Tsay, 1987]), que usa coeficientes aleatórios para produzir heteroscedasticidade condicionada, e o modelo SV (*stochastic volatility*) (ver [Taylor, 1994] e [Harvey et al., 1994]), que introduz um termo erro adicional na equação da variância condicional para captar melhor a evolução da volatilidade.

Capítulo 4

Metodologia

A necessidade de observação e análise de mais do que uma variável estatística no tempo motivou o surgimento de uma série de novas técnicas: a análise multivariada de dados (ver [Johnson & Wichern, 2007]). A necessidade de meios computacionais para implementar muitas destas técnicas e o desenvolvimento destes meios desencadearam um progresso muito acentuado desta área nos últimos tempos.

4.1 Classificação e agrupamento de séries temporais

Uma técnica muito utilizada na análise multivariada de dados é a análise de clusters. Esta consiste num conjunto de métodos estatísticos que agrupa os dados com o mesmo comportamento ou as mesmas características (algoritmos para agrupar objeto). Este processo necessita de uma medida de proximidade (ou afastamento) entre objeto. A escolha desta medida é subjetiva e gera grandes discussões (ver por exemplo [Caiado et al., 2006]).

Os algoritmos de clustering ou agrupamento podem ser classificados em dois tipos: hierárquicos e não-hierárquicos. Clustering não-hierárquico tem por base um ponto central e a distância é medida a partir desse ponto central. No clustering hi-

erárquico constrói-se uma árvore hierárquica para ver a relação entre as entidades. Os métodos não-hierárquicos são menos populares devido à dificuldade de encontrar o ponto central. Existem dois métodos de clustering hierárquico: Aglomerativo e Divisivo. O primeiro método começa com todos os objetos formando cada um deles um cluster e agrupa-os sucessivamente consoante a sua similaridade, de forma a que os objetos mais similares façam parte do mesmo cluster. O processo finaliza-se quando todos os objetos formam um único cluster. O segundo método começa com todos os objetos no mesmo cluster e separa-os sucessivamente consoante os objetos vão ficando mais afastados. No final cada objeto formará um cluster. Dentro dos métodos aglomerativos, são muito utilizados os métodos SL (*Single Linkage* ou vizinho mais próximo), CL (*Complete Linkage* ou vizinho mais afastado), AL (*Average Linkage* ou ligação média) e o método de Ward (ver [Johnson & Wichern, 2007], capítulo 12). A sua fama advém do facto de poderem ser utilizados para aglomerar tanto items como variáveis. Na técnica CL a distância entre dois clusters é definida pela distância máxima entre todos os pares de objetos, um de cada cluster. Por outro lado, a técnica SL usa a distância mínima entre todos os pares de objetos como a distância entre os clusters. Os resultados de clustering hierárquico podem ser representados através de um diagrama unidimensional, um dendrograma, também conhecido como árvore de agrupamento. Depois de formados os clusters, há que decidir com quantos clusters se vai trabalhar. Para tal existe uma série de critérios que podem auxiliar a nossa escolha, entre eles o critério Duda-Hart (ver [Duda & Hart, 1973]). O índice Duda-Hart $Je(2)/Je(1)$ é definido pelo quociente entre a soma do quadrado dos erros do grupo a ser separado em dois subgrupos ($Je(1)$) e a soma do quadrado dos erros dos grupos resultantes da separação ($Je(2)$). Pretende-se assim medir a perda de informação associada ao clustering que ocorre quando se tratam por “igual” objeto no mesmo cluster quando estes não são “iguais”. Para auxiliar a decisão, associa-se um pseudo T-quadrado. Desta forma, a regra de paragem consiste em encontrar um dos maiores $Je(2)/J(1)$ valores que corres-

pondam a um valor pseudo T-quadrado baixo, tendo valores pseudo T-quadrado seguintes muito superiores. Este critério requer o uso de cluster hierárquico pois precisa de saber qual o grupo a ser separado em cada nível hierárquico, e como é separado. O índice Duda-Hart é também local, dado que a única informação usada é relativa à separação do grupo, sendo a informação dos outros grupos irrelevante.

Um método complementar ao *Data Clustering* usado principalmente na análise explanatória de dados é o *Ordination*. Este método ordena objetos caracterizados por valores em várias variáveis, de tal forma que objetos similares fiquem próximos uns dos outros e os objetos dissimilares fiquem afastados uns dos outros (ver [Johnson & Wichern, 2007], capítulo 12). Exemplos de técnicas de ordenação: *Principal Components Analysis* (PCA), *Non-metric Multidimensional Scaling* (NMDS), *Correspondence Analysis* (CA).

Multidimensional Scaling é uma técnica de redução da dimensão com o objetivo de representar os objetos graficamente, para facilitar a sua análise no sentido de proximidade. Esta técnica de redução consiste em representar os objetos num sistema coordenado de dimensão inferior ao original no qual as distâncias entre os objetos se aproximam das respetivas medidas de proximidade. Assim, os objetos mais próximos serão os mais semelhantes, de acordo com o critério adotado. O algoritmo recebe uma matriz de distâncias e transforma-a noutra matriz de dimensão inferior escolhida previamente. As linhas desta matriz são as coordenadas dos objetos no novo sistema. Seja D a matriz de distâncias Euclidianas observada, de dimensão $n \times n$. Por *multidimensional scaling*, a matriz D é transformada numa matriz T de dimensão $n \times d$ ($d < n$). A representação d -dimensional que melhor aproxima a matriz de dissimilaridades observada (D), é dada pelos d vetores próprios de TT' com os d maiores valores próprios. Esta técnica é muitas vezes usada como auxiliar de *data clustering*. A técnica de *Multidimensional Scaling* tem a capacidade de modelar relações não lineares entre variáveis, consegue tratar dados nominais e ordinários e não necessita de normalidade multivariada. Pode assim ser apresentada

como alternativa a técnicas como *Factor Analysis* (ver [Johnson & Wichern, 2007], capítulo 12).

4.2 Medidas de distância

A escolha de uma medida de distância é um problema fundamental na classificação e agrupamento de séries temporais. Em geral, podemos considerar medidas de distância entre séries temporais baseadas nas suas próprias observações, nas suas características extraídas das séries (autocorrelação, espectro) ou nos parâmetros estimados dos modelos ajustados às séries. A medida usual é a euclidiana clássica, que mede a distância entre duas observações correspondentes no espaço. Dadas duas séries temporais $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$, a distância euclidiana entre estas duas séries é dada por

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

No entanto esta medida apresenta algumas limitações: é invariante a transformações na ordem da série, não tem em consideração a estrutura de autocorrelação (ver [Galeano & Peña, 2000]) e só é válida para séries temporais com a mesma dimensão. Surgem então alternativas a esta distância, de forma a tentar colmatar estas falhas. Estas alternativas apresentam-se na forma de métodos paramétricos e não paramétricos (ver [Galeano & Peña, 2000]).

A função de autocorrelação (FAC) de uma série temporal descreve a relação de dependência de uma série temporal y_t , medindo a associação linear entre y_t e y_{t-s} , e é definida por

$$\rho_s = \frac{Cov(y_t, y_{t-s})}{\sqrt{Var(y_t)Var(y_{t-s})}}.$$

A função de autocorrelação parcial (FACP) de uma série temporal mede a corre-

lação entre y_t e y_{t-s} eliminando o efeito das variáveis intermédias, $y_{t-1}, y_{t-2}, \dots, y_{t-s+1}$.

Dentro das medidas não paramétricas podemos destacar as seguintes:

- Galeano e Peña (ver [Galeano & Peña, 2000]) sugeriram uma medida baseada na FAC estimada. Seja $X = (x_{1,t}, \dots, x_{k,t})'$ um conjunto de séries temporais e $\hat{\rho}_i = (\hat{\rho}_{i,1}, \dots, \hat{\rho}_{i,m})$ o vetor dos coeficientes de autocorrelação estimados da série temporal i para um certo m tal que $\hat{\rho}_j \cong 0$ para $j > m$. A distância entre as séries temporais x e y é definida por

$$d_{ACF}(x, y) = \sqrt{(\hat{\rho}_x - \hat{\rho}_y)' \Omega (\hat{\rho}_x - \hat{\rho}_y)},$$

onde Ω é uma matriz de pesos.

- Quando $\Omega = I$ (onde I é a matriz identidade), obtém-se a distância Euclidiana entre os coeficientes de autocorrelação das séries temporais x e y .
- Quando $\Omega = [cov(\hat{\rho})]^{-1}$ é a matriz inversa das covariâncias das autocorrelações, obtém-se a distância de Mahalanobis entre as autocorrelações.
- Também é comum o uso de pesos que diminuem com o lag de autocorrelação, por exemplo, com um decaimento geométrico (ver [Caiado et al., 2006])

$$d_{ACFG}(x, y) = \sqrt{\sum_{i=1}^L m_i (\hat{\rho}_{i,x} - \hat{\rho}_{i,y})^2},$$

onde L é o número de autocorrelações, $m_i = pq^i$ para $i = 1, \dots, L$, $p = 1 - q$ e $0 < p < 1$.

- Caiado, Crato e Peña (ver [Caiado et al., 2006]) apresentaram uma distância baseada na função de autocorrelação parcial (FACP) e uma outra baseada

na função de autocorrelação inversa (FACI). A medida FACP entre as séries temporais x e y é definida por

$$d_{FACP}(x, y) = \sqrt{(\hat{\phi}_x - \hat{\phi}_y)' \Omega (\hat{\phi}_x - \hat{\phi}_y)}$$

onde $\hat{\phi}$ é o vetor dos coeficientes estimados da autocorrelação parcial e Ω é uma função de pesos.

Caiado e Crato (ver [Caiado & Crato, 2010]) introduziram uma distância paramétrica, baseada nos parâmetros estimados de um modelo TGARCH(1,1) ajustado às séries temporais. Seja $r_{x,t} = \log P_{x,t} - \log P_{x,t-1}$ o retorno composto contínuo de um ativo x no intervalo de tempo $[t-1, t]$ ($r_{y,t}$ é definido de forma similar para o ativo y). Suponha-se que se ajusta um modelo TGARCH(1,1) a ambas as séries pelo método da máxima verosimilhança assumindo que os erros seguem uma distribuição GED (*generalized error distribution*). Sejam $T_x = (\hat{\alpha}_x, \hat{\beta}_x, \hat{\gamma}_x, \hat{\nu}_x)'$ e $T_y = (\hat{\alpha}_y, \hat{\beta}_y, \hat{\gamma}_y, \hat{\nu}_y)'$ os vetores dos parâmetros ARCH, GARCH, *leverage effect* e *tail-thickness* estimados, e V_x e V_y as respetivas matrizes de variâncias e covariâncias estimadas. A distância entre as características da dinâmica das séries dos retornos $r_{x,t}$ e $r_{y,t}$ pode ser definida por

$$d_{TGARCH}(x, y) = \sqrt{(T_x - T_y)' \Omega^{-1} (T_x - T_y)},$$

onde $\Omega = V_x + V_y$ é a matriz de pesos.

A escolha da medida deve ter em conta os objetivos do estudo em causa. Sabendo que as autocorrelações do valor absoluto dos retornos são geralmente maiores do que as autocorrelações do quadrado dos retornos, em particular para índices bolsistas (ver [Franses & van Dijk, 2000]), as medidas usadas serão a distância baseada na FAC do quadrado dos retornos e do valor absoluto dos retornos de ativos financeiros. Esta escolha deve-se ao interesse em estudar a semelhança entre a dependência não

linear e os aparentes efeitos heterocedásticos autoregressivos (ver [Engle, 1982]) de séries financeiras. Assim, em particular, as distâncias utilizadas serão

$$d_{ACFA}(x, y) = \sqrt{(\hat{\rho}_{|r_x|} - \hat{\rho}_{|r_y|})' \Omega (\hat{\rho}_{|r_x|} - \hat{\rho}_{|r_y|})}, \quad \text{e}$$

$$d_{ACFQ}(x, y) = \sqrt{(\hat{\rho}_{r_x^2} - \hat{\rho}_{r_y^2})' \Omega (\hat{\rho}_{r_x^2} - \hat{\rho}_{r_y^2})},$$

onde $r_x = (r_{x,1}, \dots, r_{x,n})$, $r_y = (r_{y,1}, \dots, r_{y,n})$ e Ω é uma matriz de pesos.

Capítulo 5

Simulação

Pretende-se verificar se o método de agrupamento consegue classificar corretamente séries financeiras geradas pelo mesmo processo. Para tal, consideram-se dois processos diferentes e geram-se duas séries com a mesma dimensão (N) para cada um dos dois processos considerados. São calculadas as distâncias entre cada par de séries através dos métodos da FAC com ponderadores uniformemente distribuídos, $d_{FAC}(x, y)$, e geometricamente decrescentes com a antiguidade, $d_{FACG}(x, y)$. As quatro séries são agrupadas através de um algoritmo de agregação (*single linkage*, *complete linkage* ou *average linkage*). Regista-se um sucesso se as duas séries geradas pelo mesmo processo pertencerem ao mesmo cluster.

Para ilustrar o desempenho do método de agrupamento do vizinho mais afastado aplicado a séries financeiras, realizou-se um conjunto de simulações de Monte Carlo:

- (a) GARCH(1,1) com $\alpha = 0,8$ e $\beta = 0,1$ contra GARCH(1,1) com $\alpha = 0,1$ e $\beta = 0,8$
- (b) GARCH(1,1) com $\alpha = 0,8$ e $\beta = 0,1$ contra ARCH(5) com $\alpha_1 = 0,4$, $\alpha_2 = 0,25$, $\alpha_3 = 0,15$, $\alpha_4 = 0,08$ e $\alpha_5 = 0,02$
- (c) GARCH(2,1) com $\alpha = 0,15$ e $\beta_1 = 0,7$, $\beta_2 = 0,1$ contra GARCH(1,2) com $\alpha_1 = 0,7$, $\alpha_2 = 0,1$ e $\beta = 0,15$

Na escolha dos modelos procuraram-se estruturas de parâmetros muito próximas entre si de modo a dificultar a tarefa de classificação das séries temporais. No caso (a) confrontam-se dois modelos GARCH, um dominado pela parte GARCH do processo e outro pela parte ARCH. Sabendo que o modelo GARCH(p,q) é equivalente ao modelo ARCH(∞), no caso (b) compara-se o modelo GARCH dominado pela parte GARCH do processo e com ordem pequena, com o modelo ARCH com ordem superior. O caso (c) é semelhante ao caso (a) mas utiliza duas estruturas de parâmetros menos parcimoniosas do que as utilizadas no primeiro caso.

Em cada caso, simularam-se duas séries de cada modelo. Usando a FAC do quadrado dos retornos, aplicou-se o método do vizinho mais afastado (*complete linkage*) para as distâncias $d_{FAC}(x, y)$ e $d_{FACG}(x, y)$ com $p = 0,05$, às quatro séries simuladas. Caso as duas séries geradas pelo mesmo modelo ficassem no mesmo cluster contabilizava-se um sucesso. Este processo foi realizado 1000 vezes. Aplica-se o mesmo procedimento usando a FAC do valor absoluto dos retornos.

Observando a figura A.1 constata-se que as FAC do processo GARCH(1,1) com $\alpha = 0,1$ e $\beta = 0,8$ apresentam um decaimento mais lento do que as FAC do processo GARCH(1,1) com $\alpha = 0,8$ e $\beta = 0,1$. Analisando a figura A.2, observa-se que as FAC do processo GARCH(1,1) em questão apresentam um decaimento mais brusco do que as FAC do processo AR(5). Por outro lado, a análise da figura A.3 permite concluir que as FAC do processo GARCH(2,1) decaem mais lentamente do que as FAC do processo GARCH(1,2).

Tabela 5.1: Percentagens de sucesso na comparação de séries simuladas de processos: GARCH(1,1), 0,8, 0,1 contra GARCH(1,1), 0,1, 0,8

Ponderador	N	FAC dos quadrados	FAC do valor absoluto
Uniforme	200	22,7	36,5
	500	30,9	55,6
	1000	35,5	70,0
	2000	44,2	82,9
	5000	48,2	87,9
Geométrico	200	23,7	40,0
	500	33,6	61,6
	1000	38,1	74,6
	2000	46,2	86,3
	5000	46,3	90,5

Tabela 5.2: Percentagens de sucesso na comparação de séries simuladas de processos: GARCH(1,1), 0,8, 0,1 contra ARCH(5), [0,4 0,25 0,15 0,08 0,02]

Ponderador	N	FAC dos quadrados dos retornos	FAC do valor absoluto dos retornos
Uniforme	200	22,7	20,2
	500	25,6	27,2
	1000	29,1	32,6
	2000	34,8	41,0
	5000	45,4	54,5
Geométrico	200	19,1	21,2
	500	25,7	28,6
	1000	30,9	34,9
	2000	34,1	42,9
	5000	44,7	55,4

Tabela 5.3: Percentagens de sucesso na comparação de séries simuladas de processos: GARCH(2,1), [0,7 0,1], 0,15 contra GARCH(1,2), 0,15, [0,7 0,1]

Ponderador	N	FAC dos quadrados dos retornos	FAC do valor absoluto dos retornos
Uniforme	200	21,0	30,2
	500	27,7	44,4
	1000	37,9	58,6
	2000	44,1	69,4
	5000	59,3	82,6
Geométrico	200	20,6	32,0
	500	27,8	48,0
	1000	34,4	63,3
	2000	41,0	74,8
	5000	53,4	84,9

Analisando as tabelas anteriores, constata-se que, como seria de esperar, o aumento da dimensão da amostra aumenta a eficiência do método. Nos três casos, os resultados usando a FAC do módulo dos retornos são melhores do que usando a FAC do quadrado dos retornos. Relativamente à escolha do ponderador usado na medida de distância, constata-se que o ponderador geométrico é ligeiramente melhor que o ponderador uniforme. O primeiro caso foi onde o processo teve menor êxito, com uma percentagem de sucesso máxima de 90,5%, enquanto que no primeiro e no terceiro caso as percentagens de sucesso máximas foram 55,4% e 84,9%, respetivamente.

Capítulo 6

Aplicação

6.1 Informação Geral

A base de dados utilizada para esta análise consiste nas cotações diárias das empresas constituintes do índice FTSE100 à data de 1 de Maio de 2012, durante os últimos 20 anos. Este índice é calculado pelo *FTSE The Index Company*. O FTSE100 representa 81% do mercado de capitalização de toda a *London Stock Exchange* e é o mais usado como indicador do mercado do Reino Unido. Os constituintes do índice são determinados quadrimestralmente.

A base de dados inclui as empresas indicadas no Apêndice B. As empresas foram agrupadas em 8 setores consoante a sua atividade: *Basic Materials* (BM), *Consumer Goods* (CG), *Financial* (F), *Healthcare* (H), *Industrial Goods* (IG), *Services* (S), *Technology* (T) e *Utilities* (U).

Note-se que o índice é composto por 100 empresas, mas a lista é composta por 102 entradas, uma vez que no caso das empresas Royal Dutch Shell e Schroders existem duas categorias de ações.

Cada série temporal é composta pelas cotações diárias da respetiva empresa. Algumas séries temporais podem apresentar dimensões diferentes, dependendo da altura em que a respetiva empresa entrou para o mercado.

6.2 Estatísticas Descritivas

A tabela C.2 mostra as estatísticas descritivas do retorno diário, $\ln(p_t/p_{t-1})$, de cada uma das empresas, onde p_t é a cotação da empresa na *London Stock Exchange* no instante t . As estatísticas descritivas são compostas pela dimensão da série, média anualizada, desvio padrão anualizado, coeficiente de assimetria e pelo coeficiente de curtose.

A tabela C.1 mostra as estatísticas descritivas de cada um dos setores. Os setores que apresentam um maior retorno médio anualizado são *Technology* (0,1118) e *Basic Materials* (0,105), enquanto que o setor com menor retorno médio é *Financial* (0,0562). Relativamente à volatilidade anualizada, o setor com menor desvio padrão é o das *Utilities* (0,2862) e o que apresenta maior desvio padrão é o *Basic Materials* (0,4371). Note-se que o setor *Technology* apresenta uma volatilidade condicional anualizada muito próxima do setor de *Basic Materials*. À exceção do setor *Consumer Goods*, que apresenta uma assimetria positiva (0,033071), todos os setores exibem assimetrias negativas, podendo-se concluir que a distribuição dos retornos apresenta *heavy tails*. O setor com maior coeficiente de assimetria é o setor *Technology* (-1,18) e também é o setor com maior excesso de curtose (38,95). O setor que exhibe o menor excesso de curtose é o setor *Utilities*.

6.3 Estimação GARCH

A tabela 6.1 mostra a média das estimativas dos parâmetros do modelo TGARCH(1,1), $(\hat{\alpha}, \hat{\beta} \text{ e } \hat{\gamma})$, para os retornos diários de cada uma das empresas por setor, assim como o coeficiente de persistência $(\hat{\alpha} + \hat{\beta} + \frac{\hat{\gamma}}{2})$ e o impacto à chegada de más notícias $(\hat{\alpha} + \hat{\gamma})$. A estimação com erros t-sudent e GED conduziu a resultados inconsistentes para algumas das séries observadas, portanto assumiram-se erros com distribuição Gaussiana.

Tabela 6.1: Média das Estimativas dos parâmetros do modelo TGARCH(1,1) por setor

Setor	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\gamma}$	$\hat{\alpha} + \hat{\beta} + \frac{\hat{\gamma}}{2}$
<i>Basic Materials</i>	0,0469	0,0728	0,8485	0,1197	0,9318
<i>Consumer Goods</i>	0,0197	0,0647	0,9415	0,0843	0,9935
<i>Financial</i>	0,0526	0,0816	0,8842	0,1342	0,9776
<i>Healthcare</i>	0,0495	0,0511	0,9102	0,1007	0,9853
<i>Industrial Goods</i>	0,0301	0,0532	0,9319	0,0833	0,9886
<i>Services</i>	0,0449	0,0859	0,8868	0,1308	0,9746
<i>Technology</i>	0,0357	0,0440	0,9406	0,0797	0,9982
<i>Utilities</i>	0,0394	0,0537	0,9129	0,0931	0,9792

O setor que apresenta um menor impacto na variância condicional às “boas notícias” no modelo GARCH é o *Consumer Goods*, com um impacto de 1,97%, enquanto que o setor *Financial* é o que apresenta o maior impacto (5,26%). Relativamente ao impacto na variância condicional às “más notícias”, o setor *Technology* exhibe o menor com $\alpha + \gamma = 7,97\%$. Em contrapartida, o setor *Financial* apresenta o maior impacto às “más notícias” (13,42%). Note-se que o setor com menor impacto às “boas notícias” apresenta um valor $\alpha + \gamma$ relativamente próximo (8,33%) ao das *Technology*. Todos os setores apresentam uma persistência de choques à volatilidade próxima de 100%, sendo que o setor que apresenta menor persistência é o *Basic Materials*, com 93,18%, e os que apresentam maior persistência são os setores *Technology* e *Consumer Goods*.

6.4 Análise de Clusters

Para esta análise utilizaram-se as 98 séries correspondentes às empresas com um histórico de pelo menos 2 anos de observações, excluindo as empresas International Airlines Group (IAG), Polymetal Int (POLY), Evraz (EVR) e Glencore International (GLEN) com 332, 133, 127 e 249 observações, respetivamente. Calcularam-se as autocorrelações do valor absoluto dos retornos para as últimas $n = 889$ obser-

vações (número mínimo de observações da amostra das 98 séries menos uma). Em seguida, aplicaram-se as técnicas de clustering para os primeiros 30 ($\cong \sqrt{889}$) lags (dependência a curto prazo) e para os lags 31, \dots , 80 (dependência a longo prazo).

6.4.1 Dependência de curto prazo

Utilizando as autocorrelações do valor absoluto dos retornos para os primeiros 30 lags e a distância d_{ACF} com pesos uniformes, construiu-se um dendrograma e um *scaling map* a duas dimensões.

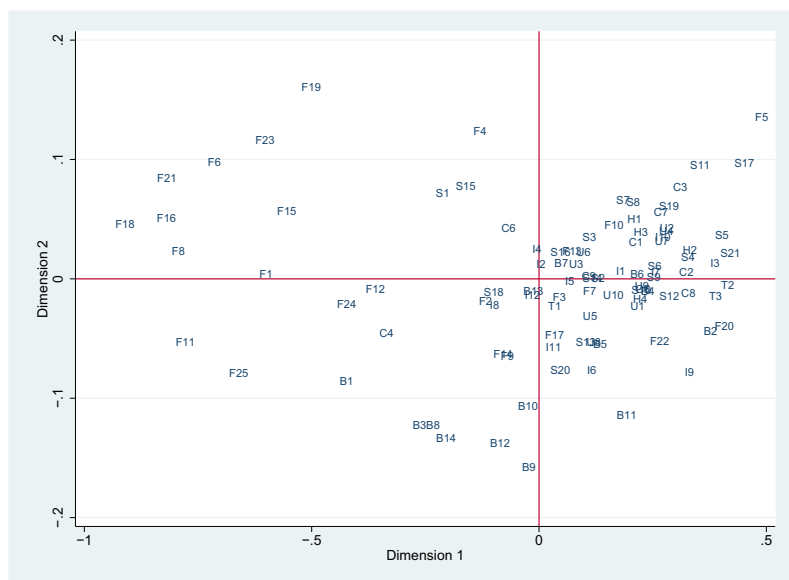


Figura 6.1: Scaling Map a duas dimensões

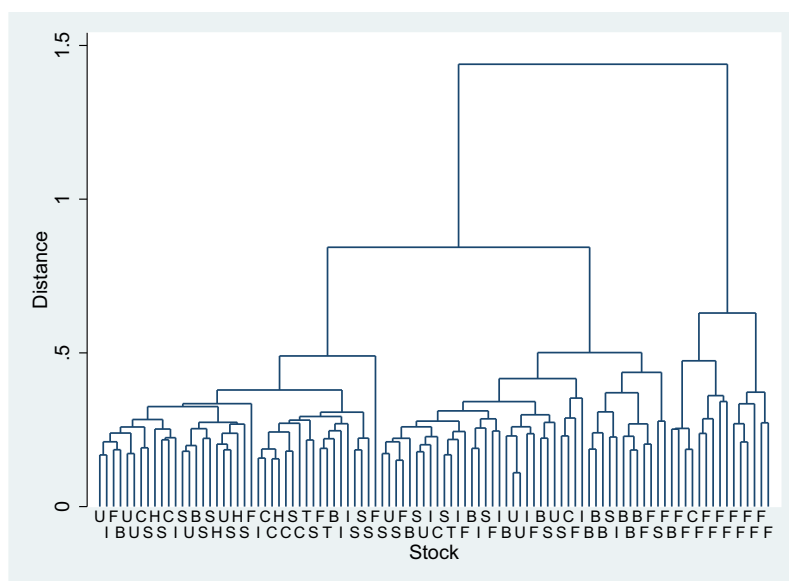


Figura 6.2: Dendrograma

Note-se que as observações estão indicadas com o respetivo setor em vez da sigla da empresa a que se refere. Pretende-se desta forma facilitar a sua análise.

De acordo com o critério Duda Hart $Je(2)/Je(1)$, dividiu-se as séries temporais em 3 clusters.

Tabela 6.2: Solução de clusters pelo método de *complete linkage*: dependência de curto prazo

Cluster 1		Cluster 2		Cluster 3	
Empresa	Setor	Empresa	Setor	Empresa	Setor
RRS	BM	JMAT	BM	XTA	BM
AMEC	BM	RIO	BM	GKN	CG
CRDA	BM	BLT	BM	SL	F
SAB	CG	AAL	BM	LLOY	F
ULVR	CG	ENRC	BM	BLND	F
ABF	CG	VED	BM	STAN	F
TATE	CG	KAZ	BM	LGEM	F
IMT	CG	PFC	BM	HSBA	F
DGE	CG	ANTO	BM	BARC	F
HL	F	FRES	BM	PRU	F
ADN	F	BATS	CG	HMSNO	F
RSL	F	BRBY	CG	OML	F
ADM	F	IAP	F	CSCG	F
SN	H	ASHM	F	LAND	F
AZN	H	EMG	F	AV	F
SHP	H	RBS	F		
GSK	H	SDRC	F		
SMIN	IG	SDR	F		
CRH	IG	RSA	F		
RB	IG	WEIR	IG		
BA	IG	RR	IG		
MKS	IG	MGGT	IG		
WTB	S	REX	IG		
CPG	S	NXT	IG		
ITRK	S	IMI	IG		
MORW	S	KGF	IG		
CPI	S	SBRY	S		
AGK	S	ITV	S		
TSCO	S	WPP	S		
EXPN	S	IHG	S		
BNZL	S	REL	S		
SERC	S	BSY	S		
PSON	S	CCL	S		
GFS	S	WOS	S		
SGE	T	VOD	T		
ARM	T	BTA	T		
SVT	U	BP	U		
TLW	U	RDSB	U		
IPR	U	SSE	U		
NG	U	RDSA	U		
UU	U	BG	U		
		CNA	U		

Constata-se que os clusters 1 e 2 não apresentam um setor predominante; o cluster 2 não apresenta ações do setor *Healthcare*; o cluster 3 é constituído predominantemente por empresas financeiras (86,67%). Os clusters 1 e 2 encontram-se relativamente próximos, enquanto que o cluster 3 está visivelmente mais afastado destes dois.

O setor de *Basic Materials* encontra-se maioritariamente no cluster 2 enquanto que o setor *Consumer Goods* encontra-se maioritariamente no cluster 1; o setor de *Healthcare* encontra-se exclusivamente no cluster 1; o setor *Services* encontra-se no cluster 1 e no cluster 2, com maior incidência no cluster 1, onde apresenta 60% de todas as empresas deste setor; assim como o setor *Services*, os setores *Utilities* e *Industrial Goods* encontram-se nos clusters 1 e 2, com maior incidência neste último; o setor *Financial* encontra-se distribuído por todos os clusters, onde metade das observações encontra-se no cluster 3 (54,17%); o setor *Technology* encontra-se igualmente distribuído pelos clusters 1 e 2.

Analisando a figura 6.1 é evidente a existência de um aglomerado de observações. As empresas que se encontram fora desse aglomerado são na sua maioria do setor *Financial*. Fixando a primeira coordenada, observa-se que a maioria das empresas encontra-se entre os valores 0 e 0,5, só os setores *Financial*, *Consumer Goods* e *Basic Materials* é que apresentam empresas fora deste intervalo; as empresas dos setores *Technology*, *Industrial Goods* e *Healthcare* estão próximas entre si. Analisando a segunda coordenada, a maioria das empresas encontram-se no intervalo $[-0,1; 0,1]$, apenas empresas dos setores *Financial* e *Basic Materials* é que se encontram fora deste intervalo, sendo que as empresas de *Financial* estão acima de 0,1 e as empresas de *Basic Materials* estão abaixo de $-0,1$.

6.4.2 Dependência de longo prazo

Utilizando as autocorrelações do valor absoluto dos retornos para os lags $n = 31, 32, \dots, 80$ e a distância d_{ACF} com pesos uniformes, construiu-se um dendro-

grama e um *scaling map* a duas dimensões.

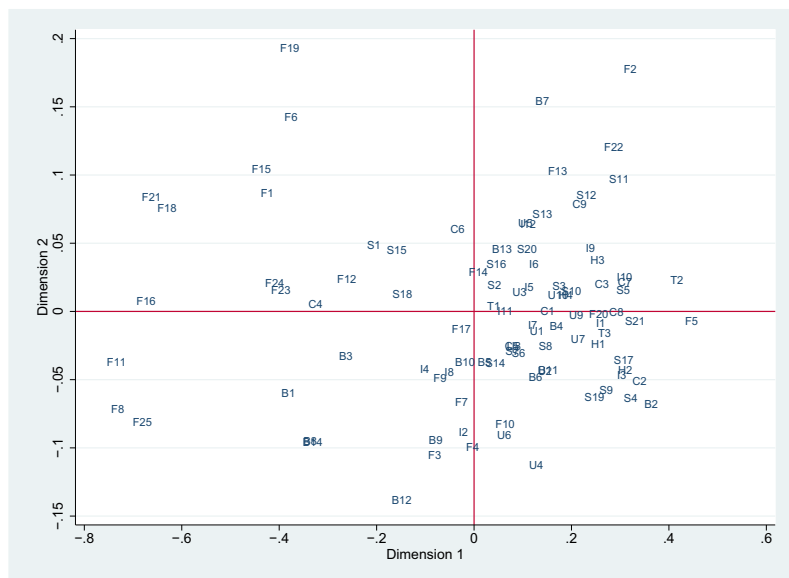


Figura 6.3: Scaling Map a duas dimensões

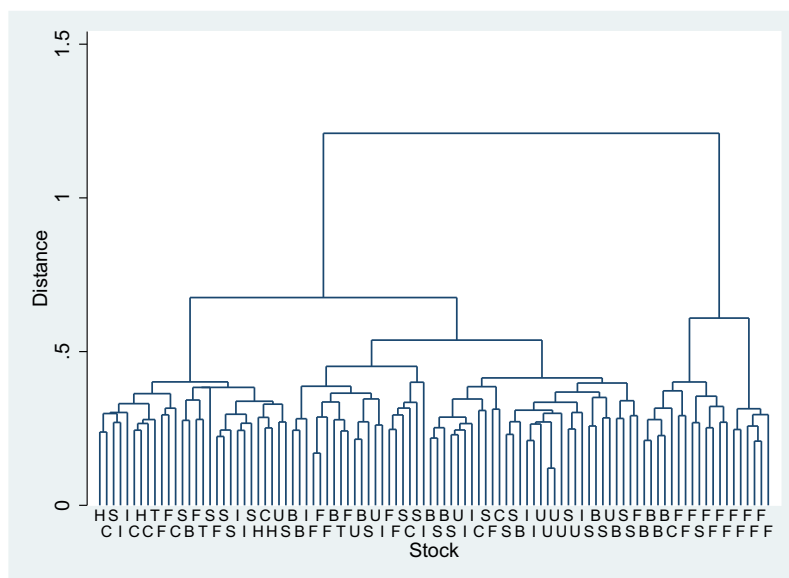


Figura 6.4: Dendrograma

De acordo com a regra de paragem Duda Hart Je(2)/Je(1), dividiu-se as séries temporais em 4 clusters.

Tabela 6.3: Solução de clusters pelo método de *complete linkage*: dependência de longo prazo

Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Empresa	Setor	Empresa	Setor	Empresa	Setor	Empresa	setor
CRDA	BM	BLT	BM	XTA	BM	LAND	F
BATS	CG	AAL	BM	KAZ	BM	AV	F
ABF	CG	JMAT	BM	RIO	BM	LGEN	F
DGE	CG	VED	BM	ENRC	BM	HMSNO	F
SAB	CG	FRES	BM	GKN	CG	BLND	F
IMT	CG	AMEC	BM	BARC	F	OML	F
HL	F	PFC	BM	LLOY	F		
ADN	F	ANTO	BM	STAN	F		
RSL	F	RRS	BM	CSCG	F		
SN	H	TATE	CG	PRU	F		
SHP	H	BRBY	CG	HSBA	F		
AZN	H	ULVR	CG	SL	F		
GSK	H	ADM	F	WOS	S		
CRH	IG	ASHM	F				
RB	IG	RBS	F				
BA	IG	IAP	F				
SMIN	IG	SDRC	F				
AGK	S	EMG	F				
CPI	S	RSA	F				
ITRK	S	SDR	F				
GFS	S	IMI	IG				
TSCO	S	NXT	IG				
WTB	S	MGGT	IG				
PSON	S	MKS	IG				
SGE	T	REX	IG				
ARM	T	KGF	IG				
BP	U	WEIR	IG				
TLW	U	RR	IG				
		SERC	S				
		BNZL	S				
		IHG	S				
		BSY	S				
		EXPN	S				
		REL	S				
		CCL	S				
		WPP	S				
		CPG	S				
		MORW	S				
		ITV	S				
		SBRY	S				
		BTA	T				
		VOD	T				
		IPR	U				
		RDSB	U				
		RDSA	U				
		NG	U				
		CNA	U				
		UU	U				
		SSE	U				
		SVT	U				
		BG	U				

À semelhança do estudo da dependência a curto prazo, constata-se que os clusters 1 e 2 não apresentam um setor predominante; o cluster 2 não apresenta ações do setor *Healthcare*; o cluster 3 só apresenta ações dos setores *Basic Materials*, *Consumer Goods*, *Services* e *Financial*, sendo que este último é predominante, com 53,85% das observações; o cluster 4 só apresenta empresas do setor *Financial*. Os clusters 1 e 2 encontram-se relativamente próximos, enquanto que o cluster 3 está visivelmente mais afastado destes dois.

O setor de *Basic Materials* encontra-se maioritariamente no cluster 2; o setor de *Healthcare* encontra-se exclusivamente no cluster 1; o setor *Industrial Goods* encontra-se maioritariamente no cluster 2, assim como o setor *Services*; o setor *Utilities* encontra-se sobretudo no cluster 2 onde este cluster tem 81,82% das empresas deste setor; o setor *Financial* encontra-se disperso por todos os clusters, enquanto que o setor *Consumer Goods* encontra-se disperso pelos clusters 1, 2 e 3.

Observando a figura 6.3, é evidente a existência de um aglomerado de observações, que se encontram relativamente afastadas das restantes. Fixando a primeira coordenada, observa-se que a grande maioria encontra-se entre os valores $-0,2$ e $0,4$, com uma maior concentração entre 0 e $0,4$; os *stocks* do setor *Financial* encontram-se muito dispersos, sendo estes os que se encontram mais afastados dos restantes; as empresas dos setores *Technology*, *Utilities* e *Healthcare* encontram-se próximas entre si. Observando a segunda coordenada, verifica-se que a grande maioria das empresas encontra-se entre $-0,1$ e $0,1$; os setores *Consumer Goods*, *Healthcare* e *Services* estão, na sua maioria, próximos de zero, os setores *Financial* e *Industrial Goods* estão relativamente compreendidos entre $-0,1$ e $0,1$, o setor *Technology* encontra-se entre $-0,05$ e $0,1$ e as empresas do setor *Basic Materials* encontram-se maioritariamente abaixo de zero e dispersas.

6.4.3 Discussão

Tal como seria de esperar, a distribuição dos retornos apresenta caudas pesadas, coeficiente de assimetria negativo e excesso de curtose superior a 10. Constata-se também que os retornos apresentam uma persistência de choques à volatilidade elevada.

No caso da dependência a curto prazo, a maioria dos setores apresenta um comportamento semelhante entre eles, com exceção do setor *Financial*. As empresas deste setor que se encontram dispersas são das áreas de *Life Insurance*, *Real Estate Investment Trust* e bancos. As restantes empresas deste setor que se encontram no aglomerado são das áreas de *Asset Managers*, *Investment Services* e *Nonlife Insurance*. Relativamente à dependência a longo prazo, nota-se um afastamento das empresas financeiras, sendo este o setor mais disperso. No entanto, ao contrário do caso da dependência a curto prazo, não se visualiza um aglomerado tão evidente. Em ambos os casos não existe relação evidente entre os clusters formados e os setores.

Capítulo 7

Conclusão

Neste trabalho final de mestrado estuda-se o comportamento de séries temporais financeiras, no âmbito das suas características estilizadas.

Quando analisadas as distâncias baseadas nos coeficientes da função de autocorrelação do valor absoluto dos retornos verifica-se que, quer através da utilização dos ponderadores uniformes (distância Euclidiana) quer através dos ponderadores geométricos, os resultados permitem distinguir claramente modelos GARCH. Acresce ainda que as distâncias obtidas através da utilização de ponderadores geométricos refletem uma maior eficiência empírica.

Por outro lado, baseando a análise nas distâncias obtidas através dos coeficientes da função de autocorrelação do quadrado dos retornos, os resultados obtidos, quando comparados com os valores obtidos com a aplicação da função do valor absoluto dos retornos, demonstram uma menor eficiência empírica.

A larga maioria das séries temporais analisadas apresenta uma distribuição dos retornos com *heavy tails*.

Relativamente à dependência a longo prazo, não são visíveis, nas séries financeiras, evidências que justifiquem a existência de um comportamento comum. Todavia, a dependência a curto prazo mostra que as séries das áreas de *Life Insurance*, *Real Estate Investment Trust* e bancária exibem, em geral, comportamentos distintos

das restantes séries do setor financeiro e dos outros setores, conforme se demonstra claramente no gráfico 6.1.

A classificação e o agrupamento de séries financeiras têm-se revelado cada vez mais complexas e úteis devido ao crescimento exponencial que os mercados financeiros estão a testemunhar. Os métodos utilizados neste trabalho têm, portanto, uma crescente aplicação na atividade financeira, na medida em que, de acordo com características previamente seleccionadas pelo investidor, é possível agrupar ativos financeiros e auxiliar a técnica de diversificação de portfólios.

Para investigação futura poderia ser interessante utilizar uma medida paramétrica, apesar de a nível computacional ser mais demorada do que a não-paramétrica. A comparação entre os métodos CL e SL também poderia trazer resultados interessantes para a análise de séries financeiras, bem como a utilização de outro critério de paragem de clusters, como por exemplo o critério Calinski-Harabasz. Relativamente à aplicação financeira, uma comparação destes resultados com resultados obtidos através da utilização das mesmas técnicas mas usando, por exemplo, o índice europeu poderia levar a outras conclusões? Será que teriam o mesmo comportamento?

Bibliografia

- [Black, 1976] Black, F. (1976). Studies in Stock Price Volatility Changes. *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, (pp. 177–181).
- [Bollerslev, 1986] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
- [Caiado & Crato, 2010] Caiado, J. & Crato, N. (2010). Identifying common dynamic features in stock returns. *Quantitative Finance*, 10(7), 797–807.
- [Caiado et al., 2006] Caiado, J., Crato, N., & Peña, D. (2006). A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis*, 50, 2668–2684.
- [Cont, 2001] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1, 223–236.
- [Duda & Hart, 1973] Duda, R. & Hart, P. (1973). *Pattern Classification and Scene Analysis*. Wiley, New york.
- [Engle, 1982] Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK inflation. *Econometrica*, 50(4), 987–1007.
- [Francq & Zakoian, 2010] Francq, C. & Zakoian, J.-M. (2010). *GARCH Models - Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, Inc.

- [Franses & van Dijk, 2000] Franses, P. H. & van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge University Press.
- [Galeano & Peña, 2000] Galeano, P. & Peña, D. (2000). Multivariate Analysis in Vector Time Series. *Resenhas, the Journal of the Institute of Mathematics and Statistics of the University of Sao Paolo*, 4, 383–403.
- [Glosten et al., 1993] Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48(5), 1779–1801.
- [Harvey et al., 1994] Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate Stochastic Variance Models. *Review of Economics Studies*, 61(2), 247–264.
- [Johnson & Wichern, 2007] Johnson, R. & Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.
- [Mandelbrot, 1963] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4), 394–419.
- [Nelson, 1991] Nelson, D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2), 347–370.
- [Nicolau, 2008] Nicolau, J. (2008). Processos Estocásticos Aplicados às Finanças. *Instituto de Economia e Gestão, Universidade Técnica de Lisboa*.
- [Taylor, 1994] Taylor, R. (1994). Modeling stochastic volatility: a review and comparative study. *Mathematical Finance*, 4, 183–204.
- [Tsay, 1987] Tsay, R. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association*, 82, 590–604.
- [Tsay, 2005] Tsay, R. (2005). *Analysis of Financial Time Series*. John Wiley & Sons, Inc.

[Zakoian & Rabemananjara, 1993] Zakoian, J.-M. & Rabemananjara (1993). Threshold ARCH Models and Asymmetries in Volatility. *Journal of Applied Econometrics*, 8, 31–49.

Apêndice A

Gráficos

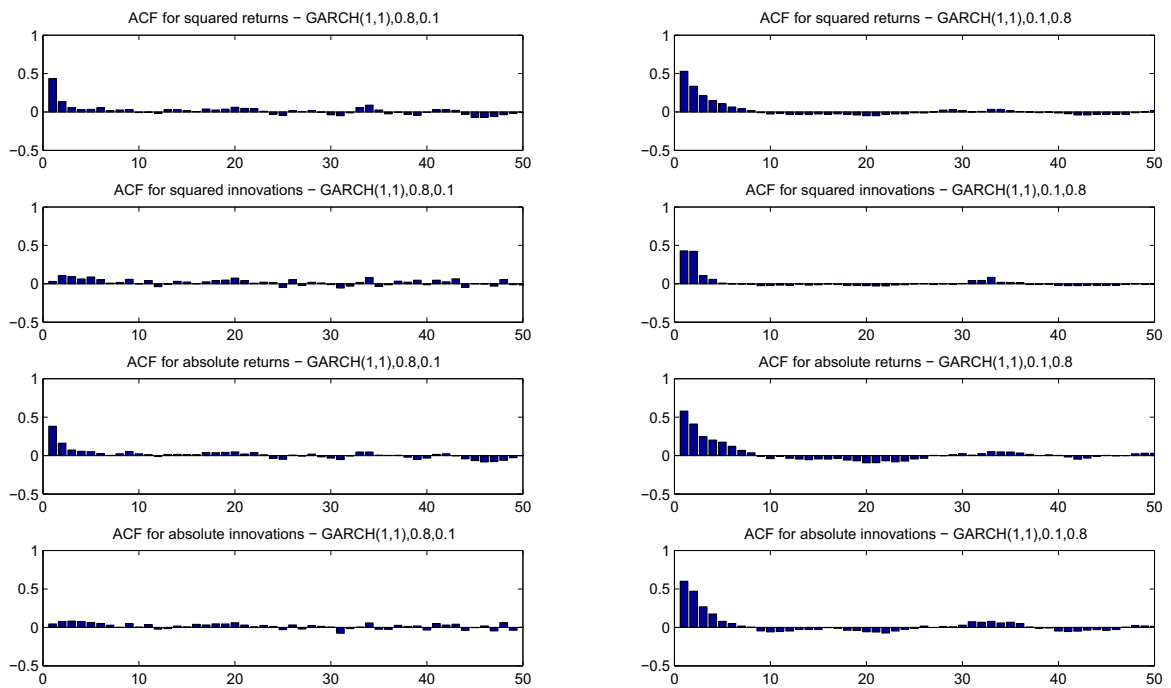


Figura A.1: Gráficos da FAC dos processos do caso (a)

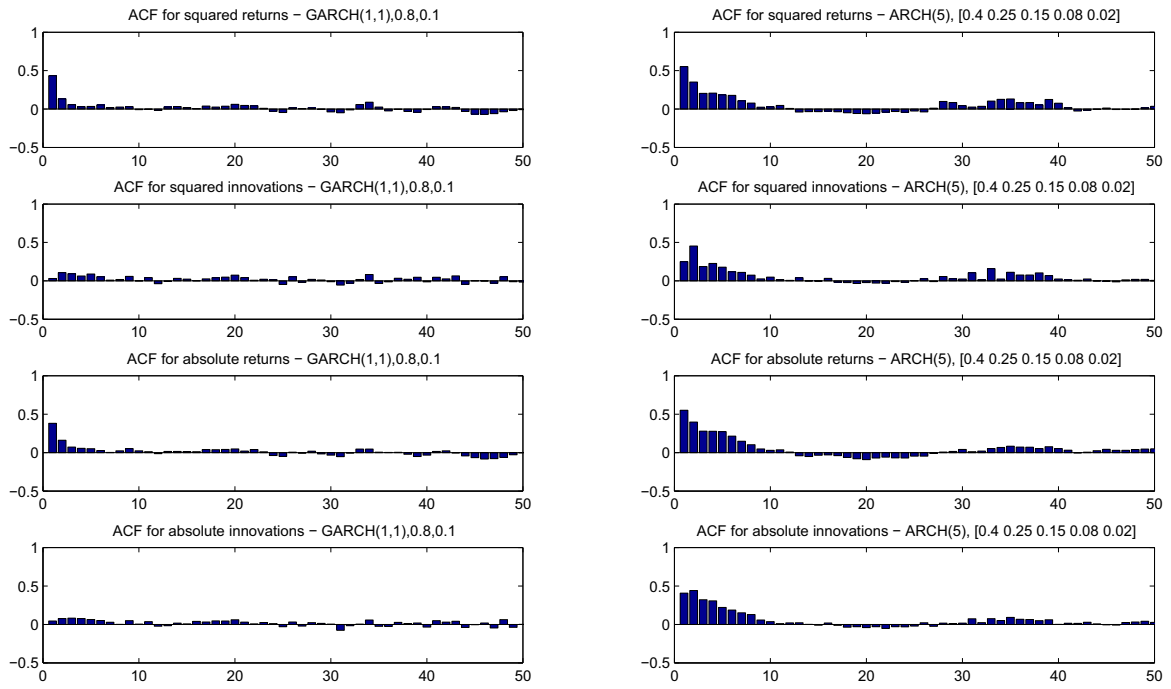


Figura A.2: Gráficos da FAC dos processos do caso (b)

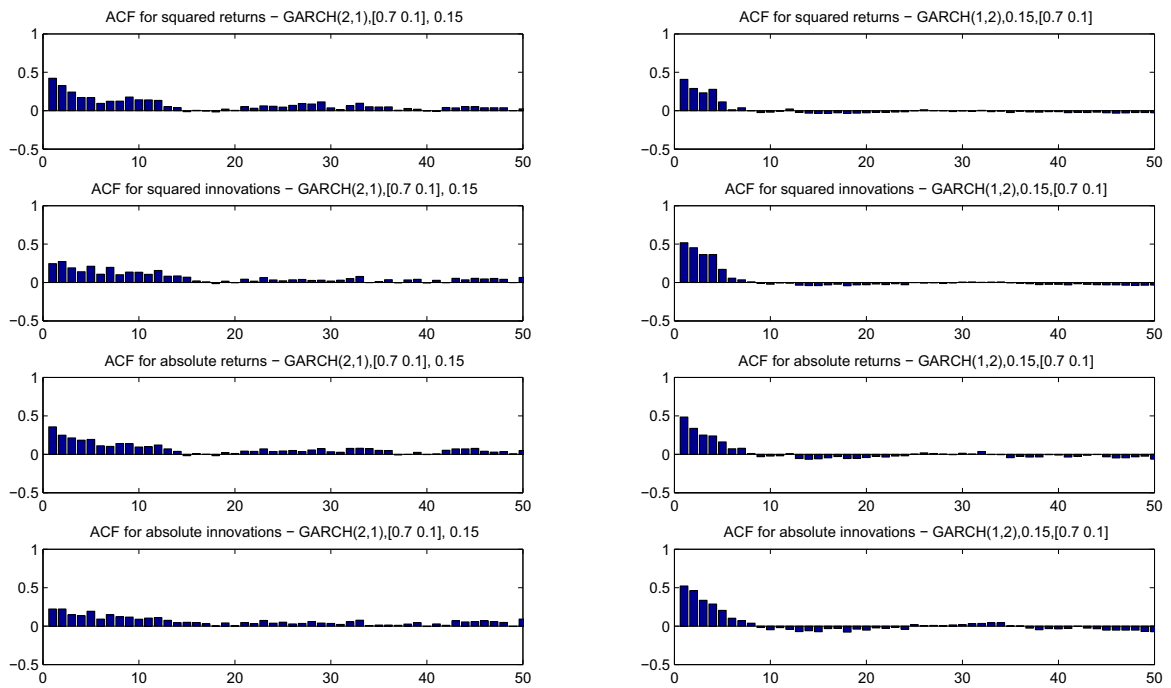


Figura A.3: Gráficos da FAC dos processos do caso (c)

Apêndice B

Empresas constituintes do índice FTSE100 a 1 de Maio de 2012

Tabela B.1: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Basic Materials*

Empresa	Sigla
AMEC	AMEC
Anglo American	AAL
Antofagasta	ANTO
BHP Billiton	BLT
Croda International	CRDA
Eurasian Natural Resources Corporation PLC	ENRC
Evrast	EVR
Fresnillo	FRES
Glencore International	GLEN
Johnson Matthey	JMAT
Kazakhmys	KAZ
Petrofac	PFC
Polymetal Int	POLY
Randgold Resources	RRS
Rio Tinto Group	RIO
Vedanta Resources	VED
Xstrada	XTA

Tabela B.2: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Healthcare*

Empresa	Sigla
AstraZeneca	AZN
GlaxoSmithKline	GSK
Shire	SHP
Smith & Nephew	SN

Tabela B.3: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Consumer Goods*

Empresa	Sigla
Associated British Foods	ABF
British American Tobacco	BATS
Burberry Group	BRBY
Diageo	DGE
GKN	GKN
Imperial Tobacco	IMT
SABMiller	SAB
Tate & Lyle	TATE
Unilever	ULVR

Tabela B.4: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Utilities*

Empresa	Sigla
BG Group	BG
BP	BP
Centrica	CNA
International Power	IPR
National Grid	NG
Royal Dutch Shell	RDSA
Royal Dutch Shell	RDSB
Scottish and Southern Energy	SSE
Severn Trent	SVT
Tullow Oil	TLW
United Utilities	UU

Tabela B.5: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Financial*

Empresa	Sigla
Aberdeen Asset Management	ADN
Admiral Group	ADM
Ashmore Group	ASHM
Aviva	AV
Barclays	BARC
Capital Shopping Centres Group	CSCG
Hammerson	HMSNO
Hargreaves Lansdown	HL
HSBC Holdings	HSBA
ICAP	IAP
Land Securities Group	LAND
Legal & General Group	LGEM
Lloyds Banking Group	LLOY
Man Group	EMG
Old Mutual	OML
Prudential	PRU
Resolution Limited	RSL
RSA Insurance Group	RSA
Schroders	SDR
Schroders NVTG	SDRC
Standard Chartered	STAN
Standard Life	SL
The British Land Company	BLND
The Royal Bank of Scotland Group	RBS

Tabela B.6: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Industrial Goods*

Empresa	Sigla
BAE Systems	BA
CRH	CRH
IMI	IMI
Kingfisher	KGF
Marks & Spencer	MKS
Meggitt	MGGT
Next	NXT
Reckitt Benckiser	RB
Rexam	REX
Rolls-Royce Group	RR
Smiths Group	SMIN
Weir Group	WEIR

Tabela B.7: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Services*

Empresa	Sigla
Aggreko	AGK
British Sky Broadcasting Group	BSY
Bunzl	BNZL
Capita Group	CPI
Carnival Corporation & plc	CCL
Compass Group	CPG
Experian	EXPN
Group 4 Securicor (G4S)	GFS
InterContinental Hotels Group	IHG
International Airlines Group	IAG
Intertek Group	ITRK
ITV	ITV
J Sainsbury	SBRY
Pearson PLC	PERSON
Reed Elsevier	REL
Serco Group	SERC
Tesco	TSCO
Whitbread	WTB
Wm Morrison Supermarkets	MORW
Wolseley	WOS
WPP	WPP

Tabela B.8: Empresas constituintes do índice FTSE100 a 1 de Maio de 2012 do setor *Technology*

Empresa	Sigla
ARM Holdings	ARM
BT Group	BTA
The Sage Group	SGE
Vodafone Group	VOD

Apêndice C

Estatísticas descritivas

Tabela C.1: Estatísticas descritivas dos retornos das empresas constituintes do FTSE100 por setor

Setor	Média	Desv.Pad.	Assimetria	Curtose
<i>Basic Materials</i>	0,1050	0,4371	-0,1970	11,7421
<i>Consumer Goods</i>	0,0992	0,2973	0,0331	12,5536
<i>Financial</i>	0,0562	0,3810	-0,4792	28,2631
<i>Healthcare</i>	0,0829	0,3002	-0,2517	14,8209
<i>Industrial Goods</i>	0,0649	0,3251	-0,4084	16,8539
<i>Services</i>	0,0688	0,3272	-0,1248	11,9986
<i>Technology</i>	0,1118	0,4246	-1,1817	38,9535
<i>Utilities</i>	0,0848	0,2862	-0,0391	10,5361

Tabela C.2: Estatísticas descritivas do retorno diário de cada uma das empresas

Stocks	Dimensão	Média	Desv.Pad	Assimetria	Curtose	Stocks	Dimensão	Média	Desv.Pad	Assimetria	Curtose
AAL	5217	0,080	0,453	-0,207	11,322	KAZ	1713	0,069	0,623	-0,221	10,465
ABF	5217	0,087	0,254	0,254	9,924	KGF	5217	0,018	0,331	-0,022	6,475
ADM	1984	0,184	0,347	-1,143	31,896	LAND	5217	0,031	0,255	-0,091	9,787
ADN	5217	0,133	0,413	-0,597	21,703	LGEN	5217	0,062	0,393	-0,207	18,060
AGK	3807	0,177	0,377	-1,553	36,183	LLOY	4263	-0,096	0,486	-0,826	36,050
AMEC	5217	0,097	0,367	0,216	12,366	MGGT	5217	0,091	0,303	0,171	10,407
ANTO	5217	0,203	0,373	0,180	11,110	MKS	5217	0,003	0,312	-0,525	19,557
ARM	3658	0,180	0,622	-4,549	127,314	MORW	5217	0,075	0,280	-0,036	11,841
ASHM	1449	0,135	0,435	-0,141	7,862	NG	4277	0,075	0,244	-0,004	11,994
AV	5217	-0,018	0,399	-0,719	23,041	NXT	5217	0,177	0,320	-0,439	15,010
AZN	4937	0,078	0,269	-0,059	8,259	OML	3342	0,017	0,441	-0,030	13,631
BA	5217	0,063	0,380	-3,406	76,619	PFC	1716	0,312	0,412	0,069	6,029
BARC	5217	0,050	0,447	1,274	43,649	POLY	133	-0,002	0,238	-0,048	4,534
BATS	5217	0,133	0,297	1,285	26,077	PRU	5217	0,059	0,405	0,024	14,258
BG	5217	0,123	0,316	0,011	6,275	PSON	5217	0,054	0,323	0,246	7,894
BLND	5217	0,057	0,310	0,120	8,596	RB	5217	0,086	0,261	-0,136	15,022
BLT	3852	0,153	0,418	0,201	7,468	RBS	5217	-0,031	0,518	-8,200	281,360
BNZL	5217	0,112	0,242	0,126	6,635	RDSA	1769	0,032	0,264	0,280	9,782
BP	5217	0,059	0,277	-0,163	8,155	RDSB	5217	0,070	0,266	0,110	7,721
BRBY	2558	0,186	0,379	-0,227	7,535	REL	5217	0,028	0,288	0,474	11,545
BSY	4539	0,055	0,346	0,192	10,986	REX	5217	0,016	0,302	-0,212	9,541
BTA	5217	-0,005	0,341	-0,329	10,249	RIO	5217	0,096	0,396	-1,115	32,661
CCL	3007	0,027	0,322	-0,994	20,441	RR	5217	0,094	0,354	0,004	10,351
CNA	3967	0,103	0,307	0,143	8,453	RRS	3871	0,160	0,486	-1,624	44,623
CPG	2933	-0,375	0,395	-0,151	3,610	RSA	5217	-0,041	0,379	-0,239	10,068
CPI	5217	0,200	0,329	0,331	10,239	RSL	888	-0,135	0,300	-0,389	10,017
CRDA	5217	0,125	0,283	0,166	10,302	SAB	3437	0,136	0,304	0,040	6,551
CRH	3342	0,015	0,367	-0,259	7,431	SBRY	5217	-0,016	0,292	-0,624	14,230
CSCG	5154	0,044	0,252	-0,049	13,797	SDR	5217	0,096	0,388	-0,019	14,303
DGE	5217	0,047	0,256	0,310	9,059	SDRC	5217	0,093	0,418	-0,422	15,284
EMG	4583	0,066	0,384	-1,516	29,620	SERC	5217	0,165	0,304	0,129	9,963
ENRC	1148	0,010	0,650	-0,229	9,463	SGE	5217	0,172	0,389	0,085	11,791
EVR	127	0,056	0,327	-0,271	3,484	SHP	4229	0,153	0,390	-1,191	32,980
EXPN	1452	0,095	0,310	0,026	7,455	SL	1517	-0,003	0,413	0,636	10,622
FRES	1038	0,267	0,585	-0,082	7,639	SMIN	5217	0,061	0,280	-0,258	10,566
GFS	4146	0,097	0,348	-0,252	18,207	SN	5217	0,066	0,272	0,079	8,495
GKN	5217	0,064	0,376	0,045	8,833	SSE	5217	0,083	0,236	-0,147	9,667
GLEN	249	-0,198	0,450	-0,025	4,070	STAN	5217	0,136	0,395	0,254	11,304
GSK	5217	0,035	0,269	0,164	9,549	SVT	5217	0,070	0,242	0,043	8,342
HL	1296	0,240	0,406	1,157	15,882	TATE	5217	0,027	0,305	-1,139	28,412
HMSNO	5217	0,030	0,281	-0,015	8,785	TLW	5217	0,233	0,428	0,486	15,248
HSBA	5168	0,090	0,302	-0,203	11,486	TSCO	5217	0,062	0,272	-0,035	7,977
IAG	332	0,080	0,417	-0,355	11,562	ULVR	5217	0,072	0,249	-0,332	9,595
IAP	3511	0,150	0,375	-0,160	17,255	UU	5217	0,027	0,239	0,020	7,633
IHG	2372	0,145	0,322	0,062	5,453	VED	2193	0,128	0,532	-0,279	7,298
IMI	5217	0,063	0,339	0,117	7,642	VOD	5217	0,100	0,348	0,067	6,460
IMT	4066	0,140	0,256	0,062	6,997	WEIR	5217	0,092	0,353	0,064	13,625
IPR	5217	0,059	0,329	-1,209	22,626	WOS	5217	0,048	0,376	-0,621	19,891
ITRK	2593	0,186	0,275	-0,350	8,321	WPP	5217	0,134	0,384	0,242	11,885
ITV	5217	0,026	0,395	0,354	9,030	WTB	5217	0,070	0,274	0,169	8,625
JMAT	5217	0,088	0,306	0,148	7,833	XTA	2640	0,142	0,531	-0,228	8,948